Molecular Breeding (2005) 16: 343–349
DOI 10.1007/s11032-005-1911-5

Short communication

Tomato SNP discovery by EST mining and resequencing

Joanne Aaron Labate* and Angela M. Baldo
*USDA, Agricultural Research Service, Plant Genetic Resources Unit, 630 W. North Street, Geneva, NY 14456, USA; *Author for correspondence (e-mail: JL265@cornell.edu)*

Received 30 September 2004; accepted in revised form 5 August 2005

Key words: Expressed sequence tag, Genetic diversity, *Lycopersicon esculentum*, Molecular marker, Single nucleotide polymorphism, SNP prediction

Abstract

Many economically important crop species are relatively depauparate in genetic diversity (e.g., soybean, peanut, tomato). DNA polymorphism within cultivated tomato has been estimated to be low based on molecular markers. Through mining of more than 148,000 public tomato expressed sequence tags (ESTs) and full-length cDNAs, we identified 764 EST clusters with potential single nucleotide polymorphisms (SNPs) among more than 15 tomato lines. By sequencing regions from 53 of these clusters in two to three lines, we discovered a wealth of nucleotide polymorphism (62 SNPs and 12 indels in 21 Unigenes), resulting in a verification rate of 27.2% (28 of 103 SNPs predicted in EST clusters were verified). We hypothesize that five regions with 1.6–13-fold more diversity relative to other tested regions are associated with introgressions from wild relatives. Identifying polymorphic, expressed genes in the tomato genome will be useful for both tomato improvement and germplasm conservation.

Estimates of intraspecific molecular genetic variation within cultivated tomato are relatively low. Mean genetic distance (Nei 1978) among nine cultivars was approximately 0.003 using RFLPs (Miller and Tanksley 1990). DNA fragment-based markers can uniquely identify tomato cultivars despite high levels of monomorphism and low polymorphism information content (PIC) values. About 50% of 129 simple sequence repeat (SSR) markers developed from expressed sequences were polymorphic among a geographically broad sample of 19 tomato cultivars (He et al. 2003). Mean PIC of the variable markers was 0.37, i.e. for a given locus there was a 37% probability that two alleles sampled at random would be different. Fingerprinting of 521 tomato varieties grown in Europe using 20 SSRs gave a mean PIC value of 0.40 (Bredemeijer et al. 2002). Among a diverse set of nine cultivars including Spanish heirloom types, 15.1% of 384 sequence related amplified polymorphism (SRAP) fragments were polymorphic (Ruiz et al. 2005). Only 9.3% of 1092 amplified fragment length polymorphism (AFLP) bands were polymorphic among 74 cultivars that included 23 pre-1970 varieties and 51 modern elite lines primarily from California (Park et al. 2004).

Single nucleotide polymorphisms (SNPs) possess desirable properties as molecular markers. Biallelism makes them easy to score in high-throughput genotyping assays. At the interspecific level, highly emphasized in tomato breeding, SNPs will exhibit less homoplasy than markers based on fragment-size (Hillis et al. 1996; Alvarez et al. 2001). SNPs can be used to saturate genetic maps in plants (Bhattramakki and Rafalski 2001). Sequencing arbitrary loci in cultivated tomato

uncovers few SNPs. In a sample of four *L. esculentum* var. *esculentum* cultivars (two heirloom and two modern types), only one SNP was discovered in approximately 7 kb of sequence (Nesbitt and Tanksley 2002). Bioinformatic prediction can increase the efficiency of SNP discovery. By mining tomato ESTs Yang et al. (2004) verified SNPs in 24 of 33 tested unigenes (72.7%) and estimated an average of 1.79 SNPs per polymorphic EST.

We have developed a data mining pipeline in PERL that screens an entire National Center for Biotechnology Information (NCBI) Unigene set (Wheeler et al. 2004) and provides an annotated list of predicted SNPs and PCR primers flanking them (Baldo et al. 2004, Huntley et al. in prep). Our pipeline subclusters and aligns each Unigene using the SEAN SNP Prediction and Display Program (Huntley 2003). The consensus sequence of each subcluster was annotated using BLASTn against sequences of mapped markers in tomato from the Solanaceae Genomics Network (SGN) (Mueller et al. 2004). SEAN, in turn, invokes Phrap (Green 2004). SEAN applies criteria designed to screen out potential sequencing errors (Picoult-Newberg et al. 1999). For a SNP to be called, there must be complete consensus among the alignment for seven nucleotides upstream of the SNP and seven downstream. Each SNP must be represented in at least two sequences. Using this method we identified 2527 potential SNPs among 764 EST clusters from the NCBI tomato Unigene set.

Using the SEAN display package, predicted polymorphisms were visually inspected within the context of lines and clones in which they were identified. Clusters interpreted to be redundant with the Yang et al. (2004) verified set based on their representative EST sequences were eliminated, then PCR primers were designed to amplify regions of predicted SNPs within 85 EST clusters. Eighty-four primer pairs amplified fragments from genomic DNA that were resequenced in two or three lines predicted to contain SNPs. Fifty-three primer pairs gave unambiguous DNA sequence data indicating whether or not SNPs were detected. The 31 remaining pairs either gave poor quality sequence, more than one PCR product, or insufficient data (line unavailable, or amplicon too large to reach predicted SNP position by direct sequencing). A total of 62 SNPs and 12 insertion–deletion (indel) polymorphisms were verified by two-pass sequencing within 21 of the 53 (39.6%) EST clusters (Table 1). A total of 31 SNPs were predicted within cDNAs for the 21 amplicons in Table 1. The single predicted SNP in amplicon 241_2 was not confirmed because, it was too close to the forward primer and an intron prevented confirming it by the reverse primer. Amplicon 2875_4 was predicted to contain two SNPs, neither of which was confirmed. This resulted in 28 cDNA SNPs in Table 1 that were predicted by SEAN and verified. Five additional cDNA SNPs reported in Table 1 were visible in the raw ESTs but not predicted, yielding a total of 33. In 32 amplicons, we did not verify any SNPs (72 predicted, unpublished results). Thirty-six of 84 amplicons apparently included introns containing 29 of the 62 SNPs plus all indels. Some of the EST cluster consensus sequences had perfect BLASTn (Altschul et al. 1990) scores (E-value equal to zero) against DNA sequences of mapped tomato markers from SGN (Mueller et al. 2004). In this way, eight polymorphic regions were virtually mapped (Table 1).

Population diversity, θ, is a measure that permits comparison of nucleotide diversity among loci and among studies because, it corrects for number of alleles sampled and size of the sequenced region. $\Theta = S/(m\Sigma\ 1/i)$ where S equals the number of segregating sites in the sample, m equals the size of the sequenced region in nucleotides, and $i = 1\ldots\ (n-1)$ with n equal to sample size (Watterson 1975). DnaSP (Rozas et al. 2003) was used to estimate θ for all sites and also separately for untranslated regions (UTRs), exons, and introns (Table 1). In the present study we assumed a sample size of two for all loci because we observed a maximum of two haplotypes, and alleles were sampled to target each haplotype. Non-zero estimates of θ ranged from 0.0015 to 0.0193 when all sites were included and from 0.0023 to 0.0235 for exons (Table 1).

Discussion

Genetic bottlenecks, founder events, and selection have contributed to the uniformity of tomato (*Lycopersicon esculentum* cv. *esculentum*) (Miller and Tanksley 1990). This lack of genetic diversity creates a challenge for characterizing crop germplasm

Table 1. Polymorphisms discovered among 21 tomato EST clusters by resequencing two or three tomato lines (23 sequenced segments).

NCBI Unigene ID and subcluster	Chr[a], Marker	GenBank ID[b]	Base pairs	Tomato Line[c]	θ[d]				Number of:				
					UTR	Exon	Intron	All	SNPs				Indels
									UTR	Exon	Intron	Total	Intron
2486_1		BV448059	259	T, E	0.0112	0.0235	na	0.0193	1	4		5	
2534_1R[e]	9, T0649	BV448061	575	T, R	na	0.0000	0.0166	0.0157			9	9	
437_2	1, 10, T0646	BV448073	662	T, R, M	na	0.0109	0.0169	0.0143		3	6	9	9
2325_3		BV448056	403	T, R	na	0.0124	na	0.0124		5		5	
220_1		BV448055	170	T, R	na	0.0118	na	0.0118		2		2	
175_1		BV448053	138	T, R, M	na	0.0073	na	0.0073		1		1	
3197_2		BV448068	147	T, R	0.0068	na	na	0.0068	1			1	
3284_1	1, T0214	BV448069	155	T, R	0.0083	0.0000	na	0.0065	1			1	
3674_2		BV448072	166	T, R, M	na	0.0060	na	0.0060		1		1	
1287_1		BV448052	172	T, R	0.0000	0.0094	na	0.0058		1		1	
3081_1		BV448065	173	T, M	na	0.0058	na	0.0058	1			1	
3332_3	7, T0643	BV448071	172	T, R	0.0069	0.0000	na	0.0058	1			1	
1909_2		BV448054	175	T, R	0.0000	0.0182	na	0.0057		1		1	
3155_3	5, cLET6I4	BV448067	843	T, R, M	0.0061	0.0050	0.0042	0.0047	1	1	2	4	
2875_4		BV448062	1,347	T, R, M	na	0.0000	0.0047	0.0045			6	6	1
3017_4		BV448064	530	T, R	na	0.0086	0.0024	0.0038		1	1	2	
2534_1F[e]	9, T0649	BV448060	601	T, R	na	0.0094	0.0000	0.0033		2		2	2
241_2R[e]		BV448058	649	T, R	na	na	0.0031	0.0031			2	2	
296_1		BV448063	975	T, R, M	na	0.0313	0.0011	0.0031		2	1	3	
3132_3		BV448066	662	T, R	na	0.0050	0.0018	0.0030	1		1	2	
1260_2	6, T0805	BV448051	416	T, R, M	na	0.0024	na	0.0024		1		1	
3300_2	4, CT188	BV448070	546	T, M	na	0.0023	0.0000	0.0018		1		1	
241_2F[e]		BV448057	659	T, R	na	0.0000	0.0017	0.0015			1	1	
Total			10,595						7	26	29	62	12

[a]Loci with high confidence BLAST scores (E value = zero) to previously published, mapped DNA marker sequence.
[b]Primers, PCR protocols, thermoprofiles, and a representative amplicon sequence are available at NCBI (http://www.ncbi.nlm.nih.gov).
[c]Tomato lines sequenced and seed sources were T = TA496 (Tanksley), E = E6203 (synonymous to TA209, Tanksley), R = Rio Grande (PI 303784), and M = Moneymaker (PI 286255).
[d]Population diversity, $4N_eu$ = 4 x effective population size x mutation rate, based on the number of segregating sites between the two haplotypes. UTRs, exons, and introns within amplicons were inferred based on matches to ESTs and open reading frames. Not applicable, na, indicates such region was not present within an amplicon.
[e]Sequence of amplicon consists of two non-overlapping, forward (F) and reverse (R) segments.

collections and for continued improvement of cultivars.

The parameter θ was estimated to be 0.0019 at locus TG10 (anonymous genomic sequence on chromosome 9) and zero at loci *fw2.2* 5′ UTR, *Adh2*, and TG11 (anonymous genomic sequence on chromosome 10) in a sample of four *L. esculentum* var. *esculentum* cultivars (Nesbitt and Tanksley 2002), i.e., between zero and 1.9 polymorphic sites per kb of sequence.

Our observation of 74 polymorphisms from resequencing approximately 20 kb (53 amplicons) in each of two to three tomato lines illustrates the utility of EST databases for computationally-aided SNP discovery (Gupta and Rustgi 2004). There are three notable observations from our results. First, SNP prediction from ESTs combined with resequencing can provide a wealth of SNPs for DNA marker development within *L. esculentum* var. *esculentum*. Theta values for 18 of the sequences in Table 1 ranged from 1.5 to 7.3 SNPs per kb. These values were similar to random diversity found within *L. esculentum* var. *cerasiforme* (1.8–5.4 SNPs per kb, Nesbitt and Tanksley 2002). Exons in Table 1 were generally as polymorphic as UTRs and introns. This result seems counterintuitive because, the principle that noncoding regions evolve more rapidly than coding regions (Hartl and Clark 1989) has been observed in tomato (Nesbitt and Tanksley 2002). For functional sequences evolutionary rates generally vary among sites (Li 1997). Clustering of SNPs within a gene has been observed in several plant studies (Huttley et al. 1997; Kawabe et al. 2000; Délye et al. 2004). The most parsimonious interpretation of inflated polymorphism within exons in the present study reflects the fact that primers were designed to target a cDNA SNP within a preferentially small (200–400 bp) amplicon.

Second, these data lend preliminary support to the hypothesis that genetic variation in domesticated tomato is unevenly distributed, with rare islands of polymorphism that originated from introgression (van der Beek et al. 1992). In the early 1940s, closely related wild species within the genus *Lycopersicon* were used as sources of disease resistance, and provided much of the breeding germplasm during subsequent decades (Stevens and Rick 1986). In general, introgression events in tomato have been documented, but the extent and persistence of linkage drag are not well known (van der Beek et al. 1992). Linkage drag associated with a gene in a self-fertilizing species bred by backcrossing, such as tomato, has been predicted to encompass 5 cM in both directions after 20 generations (Young et al. 1988 and references therein). In two tomato breeding lines, introgressions containing *Cf-ECP2* were estimated to be as large as 26 and 33 cM (Haanstra et al. 1999). Five of the sequences in Table 1 (2486_1, 2534_1, 437_2, 2325_3, and 220_1) were approximately 1.6–13-fold more diverse overall and 3.5–15-fold more diverse in introns relative to the other 18. One of these highly polymorphic regions, 437_2, contained 9 of the 12 discovered indels. With sample sizes of $n = 2$ alleles, we lack statistical power to confirm that θ estimates among loci in Table 1 are significantly different from each other. Standard deviations of θ are high, approximately equal to θ values themselves (results not shown). However, there appear to be two classes of polymorphism values (0.0015–0.0073 vs. 0.0118–0.0193), including results from introns (0.0011–0.0047 vs. 0.0166 to 0.0169) in the data. The 0.0015–0.0073 range corresponds to θ estimates that have been observed within *L. esculentum* (0.0016–0.0054, Nesbitt and Tanksley 2002). We hypothesize that the five most polymorphic regions in Table 1 ($\theta = 0.0118$ to 0.0193) represent introgressions. Sampled accessions of Rio Grande (PI 303784) and Moneymaker (PI 286255) were collected in the early 1960s and are expected to contain fewer introgressions than modern lines TA496 and E6203. E6203 (synonymous with FM6203) contains at least two introgressions in its pedigree (Court Nichols, personal communication, 2005), *Ve* on chromosome 9 (Zamir et al. 1993) from Peru Wild (Kawchuk et al. 1998), and *I* on the short arm of chromosome 11 from *L. pimpinellifolium* (Ori et al. 1997). Tobacco Mosaic Virus resistance gene *Tm-2*^{a} originated in *L. peruvianum* (Young et al. 1988), and line TA496 was developed by introgressing *Tm-2*^{a} into E6203 (synonymous with TA209) from Vendor-Tm2a (Tanksley et al. 1998, Yates et al. 2004). The most polymorphic region (2486_1) showed five cDNA SNPs that should have directly resulted from this cross. The second most polymorphic region, 2534_1R, perfectly matched COS marker T0649, which falls within approximately 0.3 cM of RFLP marker TG101 (Tanksley et al. 1992). Introgressed *Tm-2*^{a} has been mapped within 0.4 ± 0.4 cM of RFLP marker TG101 (Young

et al. 1988). Region 2534_1F, which is predicted to be highly polymorphic because of its tight linkage to 2534_1R, showed evidence of two additional SNPs and a five bp indel that were not scored because they fell too far from both primers to obtain high quality reads. Another hypothesized introgressed region, 437_2, matched COS marker T0646, which maps to chromosomes 1 and 10. Two map positions could indicate that differences between paralogs were observed. If this were the case, one paralog amplified preferentially in TA496 and the other in Rio Grande and Moneymaker, or more likely the paralogs are highly conserved and were simultaneously amplified. Although there are several clusters of introgressed *Cf* disease resistance genes on chromosome 1, known *Cf* map locations are not within 50 cM of T0646 (van der Beek et al. 1992; Haanstra et al. 1999, 2000). Sequencing homologs from wild relatives can address whether the five hypothetically introgressed alleles within TA496 are more closely related to alleles within wild species than to other *L. esculentum* alleles.

Third, SNP confirmation within cDNAs was "all or none". In the 21 EST clusters containing a verified SNP (Table 1) all other SNPs predicted by SEAN were verified, and five unpredicted SNPs that were visible in the ESTs were verified. Sequencing error, clustering of paralogous ESTs, or within-line heterogeneity may explain why 32 of 53 amplicons (60%) did not confirm predicted SNPs. Within-line heterogeneity was observed in SSR and SRAP fingerprinting (Bredemeijer et al. 2002; He et al. 2003; Ruiz et al. 2005). In examining raw EST data, tomato lines sometimes appear to be heterogeneous at polymorphic sites. Additional sampling within and among lines can address this hypothesis.

SEAN SNP prediction could possibly be improved through incorporating additional information from ESTs regarding linkage disequilibrium (expected to be high within loci), within-line homozygosity (expected to be high in tomato), number of EST copies per line and total number of copies, whether or not a predicted SNP is part of a "run" of identical nucleotides or near the end of a read (potential sequencing error), quality scores for trace files, and library and clone source (potential library and cloning artifacts).

The computational approach briefly introduced here predicted many additional SNPs (2527 vs. 101) compared to Yang et al. (2004) by utilizing data from approximately 148,000 (from more than 15 lines) vs. 138,000 public tomato sequences (from two lines) and applying a different set of criteria. The two methods shared some predictions. We did not test 10 EST clusters with predicted SNPs because, they were already verified (Yang et al. 2004). We confirmed SNPs in three clusters reported by Yang et al. (2004) as "not verified" or "no enzyme" (3284_1, 241_2, 296_1 corresponding to LEOH12, LEOH39, LEOH50) and in two clusters where Yang et al. (2004) verified a subset of predicted SNPs (2325_3 and 2534_1 corresponding to LEOH17 and LEOH25). No predicted SNPs were verified by either study for two other regions in common (our unpublished results, LEOH24, LEOH51). The SNP prediction rate over our 6.43×10^{6} bp of computationally analyzed tomato consensus sequences was 3.93×10^{-4}. Empirically we confirmed 28 of 103 predicted SNPs, yielding a transcriptome-wide estimate of 1.05×10^{-4} SNPs per nucleotide, i.e., 1 SNP per 9542 nucleotides. Based on resequencing results 28/103 (27.2%) of 2527 SNPs, or 21/85 (24.7%) of 764 EST clusters can be expected to yield positive results with additional testing.

Acknowledgements

The authors thank David Francis, Martha Mutschler, Larry Robertson, David Spooner, Lukas Mueller, and Steve Tanksley for helpful discussions about genetic diversity in cultivated tomato. Wencai Yang kindly provided FASTA files of published EST marker sequences. We thank Steve Tanksley for samples of seed from TA496 and TA209. Seed of PI 303784 and PI 286255 was provided by the USDA-ARS Plant Genetic Resources Unit, Geneva, NY. Robert Ahrens of the Solanaceae Genomics Network provided the sequences of mapped tomato markers. We are grateful to Derek Huntley for continued collaboration in refining SNP prediction methods, and Susan Sheffer, Katie Timmer, Warren Lamboy, and Michael D'Amico for their excellent technical support. Two anonymous reviewers made suggestions to greatly improve this manuscript. This work was funded by USDA-ARS Project Number: 1907-21000-016-00.

References

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–10.

Alvarez A.E., van de Wiel C.C.M., Smulders M.J.M. and Vosman B. 2001. Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theor. Appl. Genet. 103: 1283–1292.

Baldo A.M., Labate J.A. and Robertson L.D. 2004. Prediction of single nucleotide polymorphisms in tomato: How useful is EST sequence diversity? 12th International conference on Intelligent Systems for Molecular Biology. (ISMB 2004), 3rd European Conference on Computational Biology (ECCB 2004). Glasgow, Scotland. July 31. International Society for Computational Biology.

Bhattramakki D. and Rafalski A. 2001. Discovery and application of single nucleotide polymorphism markers in plants. In: Henry R.J. (ed.), Plant Genotyping: the DNA Fingerprinting of Plants. CABI Publishing, Oxon, UK, pp. 179–191.

Bredemeijer G.M.M., Cooke R.J., Ganal M.W., Peeters R., Isaac P., Noordijk Y., Rendell S., Jackson J., Röder M.S., Wendehake K., Dijcke M., Amelaine M., Wickaert V., Bertrand L. and Vosman B. 2002. Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor. Appl. Genet. 105: 1019–1026.

Délye C., Straub C., Michel S. and Le Corre V. 2004. Nucleotide variability at the acetyl coenzyme A carboxylase gene and the signature of herbicide selection in the grass weed *Alopecurus myosuroides* (Huds.). Mol. Biol. Evol. 21: 884–892.

Green P. 2004. Phred, Phrap, Consed. [Online] http://www.phrap.org/phredphrapconsed.html.

Gupta P.K. and Rustgi S. 2004. Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct. Integr. Genomics 4: 139–162.

Haanstra J.P.W., Laugé R., Meijer-Dekens F., Bonnema G., de Wit P.J.G.M. and Lindhout P. 1999. The *Cf-ECP2* gene is linked to, but not part of, the *Cf-4*/*Cf-9* cluster on the short arm of chromosome 1 in tomato. Mol. Gen. Genet. 262: 839–845.

Haanstra J.P.W., Thomas C.M., Jones J.D.G. and Lindhout P. 2000. Dispersion of the *Cf-4* disease resistance gene in *Lycopersicon* germplasm. Heredity 85: 266–270.

Hartl D. and Clark A. 1989. Principles of Population Genetics. 2nd edn. Sinauer Associates, Inc, Sunderland, MA.

He C., Poysa V. and Yu K. 2003. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. Theor. Appl. Genet. 106: 363–373.

Hillis D.M., Mable B.K. and Moritz C. 1996. Applications of molecular systematics: the state of the field and a look to the future. In: Hillis D.M., Moritz C and Mable B.K. (eds), Molecular Systematics. 2nd edn. Sinauer Associates, Inc, Sunderland, MA, pp. 321–381.

Huntley D. 2003. SEAN SNP prediction and display programs. [Online] http://zebrafish.doc.ic.ac.uk/SEAN.

Huttley G.A., Durbin M.L., Glover D.E. and Clegg M.T. 1997. Nucleotide polymorphism in the chalcone synthase-A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. Mol. Ecol. 6: 549–558.

Kawabe A., Yamane K. and Miyashita N.T. 2000. DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. Genetics 156: 1339–1347.

Kawchuk L.M., Hachey J. and Lynch D.R. 1998. Development of sequence characterized DNA markers linked to a dominant verticillium wilt resistance gene in tomato. Genome 41: 91–95.

Li W.-H. 1997. Molecular Evolution. 1st edn. Sinauer Associates Inc, Sunderland, MA.

Miller J.C. and Tanksley S.D. 1990. RFLP Analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. 80: 437–448.

Mueller L.A., Solow T., Taylor N., Skwarecki B. and Tanksley S.D. 2004. Solanaceae Genomics Network [Online]. http://www.sgn.cornell.edu.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590.

Nesbitt T.C. and Tanksley S.D. 2002. Comparative sequencing in the genus *Lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162: 365–379.

Ori N., Eshed Y., Paran I., Presting G., Aviv D., Tanksley S., Zamir D. and Fluhr R. 1997. The *I2C* family from the wilt disease resistance locus *I2* belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. Plant Cell 9: 521–532.

Park Y.H., West M.A.L. and St. Clair D.A. 2004. Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L.). Genome 47: 510–518.

Picoult-Newberg L., Ideker T.E., Pohl M.G., Taylor S.L., Donaldson M.A., Nickerson D.A. and Boyce-Jacino M. 1999. Milling SNPs from EST databases. Genome Res. 9: 167–174.

Rozas J., Sánchez-DelBarrio J.C., Messeguer X. and Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496–2497.

Ruiz J.J., García-Martínez S., Picó B., Gao M. and Quiros C. 2005. Genetic variability and relationship of closely related Spanish traditional cultivars of tomato as detected by SRAP and SSR markers. J. Am. Soc. Hort. Sci. 130: 88–94.

Stevens M.A. and Rick C.M. 1986. Chapter 2. Genetics and breeding. In: Atherton J. and Rudich J. (eds), The Tomato Crop. Chapman and Hall, NY, NY, pp. 35–109.

Tanksley S.D., Bernachi D., Beck-Bunn T., Emmatty D., Eshed Y., Inai S., Lopez J., Petiard V., Sayama H., Uhlig J. and Zamir D. 1998. Yield and quality evaluations on a pair of processing tomato lines nearly isogenic for the *Tm2a* gene for resistance to the tobacco mosaic virus. Euphytica 99: 77–83.

Tanksley S.D., Ganal M.W., Prince J.P., de Vicente M.C., Bonierbale M.W., Broun P., Fulton T.M., Giovannoni J.J., Grandillo S., Martin G.B., Messeguer R., Miller J.C., Miller L., Paterson A.H., Pineda O., Röder M.S., Wing R.A., Wu W. and Young N.D. 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141–1160.

van der Beek J.G., Verkerk R., Zabel P. and Lindhout P. 1992. Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. Theor. Appl. Genet. 84: 106–112.

Watterson G.A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. 7: 256–276.

Wheeler D.L., Church D.M., Edgar R., Federhen S., Helmberg W., Madden T.L., Pontius J.U., Schuler G.D., Schriml L.M., Sequeira E., Suzek T.O., Tatusova T.A. and Wagner L. 2004. Database resources of the National Center for Biotechnology Information: update. Nucl. Acids. Res. 32: D35–D40.

Yang W.C., Bai X.D., Kabelka E., Eaton C., Kamoun S., van der Knaap E. and Francis D. 2004. Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. Mol. Breeding 14: 21–34.

Yates H.E., Frary A., Doganlar S., Frampton A., Eannetta N.T., Uhlig J. and Tanksley S.D. 2004. Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. Euphytica 135: 283–296.

Young N.D., Zamir D., Ganal M.W. and Tanksley S.D. 1988. Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato. Genetics 120: 579–585.

Zamir D., Bolkan H., Juvik J.A., Watterson J.C. and Tanksley S.D. 1993. New evidence for placement of *Ve* – the gene for resistance to Verticillium race 1. Report of the Tomato Genetics Cooperative 43: 51.